Progress Report

DOE award number: DE-FG02-12ER26064/DE-SC0007341
Name of the recipient: University of Virginia
Project title: Terabit-Scale Hybrid Networking
Principal investigator: Malathi Veeraraghavan
Date of report: Nov 10, 2012
Period covered by the report: Jan. 15, 2012 - Nov. 10, 2012

A brief description of the progress/accomplishments during the current funding period and plans for the next year funding period.

Milestones	Objectives	Accomplishments
May 15, 2012	Run experiments on ESnet or ANI testbed to capture control-connection packets for file transfer applications (Section 7.1)	Gained an improved understanding of file transfer applications based on experiments executed on the DOE LIMAN testbed and the DOE ANI 100G wide-area testbed for use in HNTES design
July 15, 2012	Design of new online flow detection algorithms (Section 7.1)	Detailed study on online schemes using data analysis and engineering considerations completed
Jan 14, 2013	HNTES 3.0 prototyping and testing (Section 7.2)	Underway (to be completed by Jan. 14, 2013, Year 1 end)

Table 1: Planned Year 1 Milestones from the Proposal

A discussion of what was accomplished under these goals during this reporting period, including major activities, significant results, major findings or conclusions, key outcomes or other achievements.

Activities:
In this section, only the activities undertaken are described. Our findings from these activities are described in the next section titled "Significant results, major findings, key outcomes and achievements."

1. **DOE ANI testbed experiments:** We carried out two sets of experiments as described below. Our findings from these experiments are presented in the next section.

   - The first set of experiments consisted of a study of the influence of parameters used in the invocation of a file transfer application, a GridFTP client, on the nature of TCP flows created. Specifically, the GridFTP client, `globus-url-copy`, was initiated with different combinations of parameters such as `-r`, `-fast`, `-p`, etc. The `-r` option is used to move a whole directory of files; the `-fast` option is used to support extended mode (MODE E) in which the same data-plane TCP connections are used for multiple files; the `-p` option is used to create parallel TCP flows. The `diskpt` nodes at NERSC and ANL were used on the 100G testbed for these experiments. Control-plane packets were captured and analyzed to check if they were encrypted or not.
   - A second set of experiments were executed on the LIMAN testbed. This testbed offered us the opportunity to login to the Juniper routers and experiment with various configurations. The goals of these experiments were to gain an understanding of the adverse effects of high-rate α-flows on general-purpose flows, and to quantify the effects of different QoS mechanisms supported by the routers. These results are required for the HNTES 3.0 prototyping work item in the milestones table listed above. The `diskpt` nodes on the LIMAN testbed were used for these experiments. Figure 1 shows the experimental setup. The IP routers, named `newy-tb-rt1-1` and `bnl-tb-rt-2`, are Juniper MX routers running JunOS 10.2. The `diskpt` nodes are high-performance Linux hosts.

The link between the routers and the links to the `diskpts` are 10 Gb/s Ethernet (10 GbE), while the interfaces to the `app` hosts are 1 GbE. The application used is `nuttcp`. Three flows are created all destined to `newy-diskpt-1`: (i) a fixed-rate (3 Gbps) `nuttcp` UDP flow from `bnl-diskpt-1`, (ii) a `nuttcp` TCP flow from `bnl-diskpt2`, and (iii) a stream of repeated `pings` from `bnl-app`.

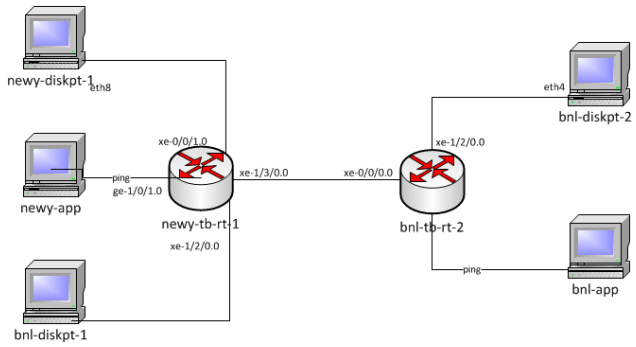

Figure 1: DOE ANI LIMAN testbed experimental setup

2. **GridFTP log analysis to determine feasibility of using an online HNTES:** GridFTP servers collect usage statistics that record the size, time, duration and other information about each transfer served. These logs served as an invaluable resource to determine whether scientific file-transfer (high-rate) flows last long enough to allow for the setup of online virtual circuits (i.e., when circuit setup action is initiated after identifying an α flow at an ingress router) given that current OSCARS IDC circuit setup delay is on the order of 1 min. In other words, if transfer times are only a small multiple of setup delay or even shorter than setup delay, then the usage of circuits is an inefficient solution. With permission from NERSC (due to privacy considerations), Globus provided us NERSC GridFTP transfer logs for one month, Sep. 2010. Similarly, we collected logs from NCAR for a three-year period (2009-2011), and with login access on the NERSC and SLAC GridFTP servers, we were able to download usage statistics logs directly from these servers (with permission from NERSC and SLAC personnel). From these logs, information about transfers on four specific paths were isolated: NCAR-NICS (2009-2011), SLAC-BNL (Feb 10 - Apr 26, 2012), NERSC-ORNL (Sept 2010), and NERSC-ANL (Mar 4 - Apr 22, 2012). Our analysis focused on characterizing the size, duration, and throughput of GridFTP transfers in production environments, to determine the suitability of using dynamic virtual circuits (VCs) in an online HNTES. Shell scripts, `awk` scripts, and R programs were implemented to extract information from the raw GridFTP transfer logs. The output files from these programs were imported into a PostgreSQL database for easier filtering and analysis. As scientists often move large numbers of files within "sessions" rather than as single file transfers, we developed a heuristic and implemented an R program to group transfers into sessions from the log files, and focussed on size and duration of sessions, not individual file transfers. For throughput, we continued using the per-transfer measure because session throughputs could be lower if some of the individual transfers within a session had lower throughput, and we are interested in finding the maximum throughput achieved by these data transfers. While single file transfers are typically short, session sizes are much larger making the overhead of dynamic VC setup negligible even under a high throughput assumption, e.g., the third-quartile of observed transfer throughput. Finally, we analyzed the SLAC-BNL GridFTP logs to determine the split between the number of transfers that use multiple parallel TCP flows vs. those that use a single TCP flow. Our findings from this activity are reported in the next section.

3. **NetFlow analysis to determine the need for an online HNTES:** In the proposal, we noted that an analysis of NetFlow data could reveal whether a large number of new α flows are identified each day. The more skewed this distribution is to the left (long tail on the left), the greater the need for an online HNTES. A flow was defined to be an α flow if it generated more than 1 GB in 1 minute, where 1 minute is denoted an α interval. We implemented an offline analysis tool (OFAT) and provided this tool to ESnet. Chris Tracy, ESnet, our collaborator executed the OFAT code against seven months' Netflow data (May - Nov. 2011) collected at 4 routers. The four routers are the BNL and NERSC provider edge (PE) routers, Sunnyvale core router, and the Equinix router at San Jose. ESnet ran scripts to

anonymize the IP addresses before providing UVA the data for further analysis. In this analysis, the methodology used was to aggregate α NetFlow reports corresponding to a given source-destination pair (/32 or /24 IP address prefixes) on a per-day (aggregation interval was set to 1 day) basis to create α prefix flows. The number of bytes in each α prefix flow is referred to as α bytes. The total time within each aggregation interval in which at least one of the constituent α flows of an α prefix flow experienced an α-interval is referred to as α time of that prefix flow. Our findings from this analysis are presented in the next section.

4. **HNTES 3.0 prototyping: Integration of OFAT and IDC client modules:** An IDC client provided by ESnet as a template for communication with the OSCARS IDC, written in Java, was downloaded and modified for integration with HNTES. It is being tested against a Test IDC system that was made available by ESnet. Much effort was required to solve authentication issues between the IDC client executed on a UVA server and the Test IDC hosted on an ESnet server. Our next step is to implement a shell script that will integrate this IDC client with OFAT to create a HNTES 3.0 in which a circuit-reservation request is sent to the OSCARS IDC for circuit setup and ingress router configuration for a Layer-3 circuit. Since OFAT identifies the IP address prefixes (/24 or /32) of α prefix flows, the ingress router will be configured to filter out packets whose source/destination IP addresses match any of these previously identified α prefix IDs, and redirect them to circuits corresponding to the appropriate egress router. This software will be tested in the next month.

5. **Offline algorithm version 2:** Based on our SLAC-BNL GridFTP log analysis, we found that a high percentage of transfers use multiple TCP streams. The implication is that our current OFAT algorithm, which requires the 1 GB threshold to be crossed in any single NetFlow report, may be too conservative in identifying α flows because while no single TCP stream may cross the 1GB threshold, the aggregate size for 8 simultaneous flows may exceed this threshold. Therefore, we have developed an algorithm that would execute the following steps: (i) prefilter out NetFlow reports with small bytes-per-packet and packets-per-NetFlow report counts to reduce processing time since these reports are seldom from α flows; (ii) use time binning to find all NetFlow reports in which one or more packets are observed within each discrete time interval; (iii) group NetFlow reports in each time bin by prefix IDs; (iv) filter these groups on total bytes to determine α prefix IDs.

6. **Online algorithm design:** Three approaches were presented in the proposal for online HNTES: port mirroring, payload analysis, real-time NetFlow data analysis. The first is infeasible at high speeds as large high-powered clusters will be required to handle packets arriving at routers on 100 Gbps links. The payload based scheme may be feasible for some applications, but GridFTP uses encrypted packets on their control-plane connections making this scheme infeasible (as described in our findings from the first set of experiments). As for real-time analysis of NetFlow data, while the Alcatel-Lucent routers support 10-sec reporting, this solution requires expensive interface cards. Besides these practical difficulties, our findings from NetFlow data analysis (presented in the next section) show that offline HNTES is highly effective in redirecting a large percentage of α bytes, making it less important to design an online HNTES.

   We studied the possible use of machine learning algorithms to determine if packet-arrival patterns for long multi-transfers sessions are different from those of short-lived single transfers as described in the GridFTP log analysis. While this work is of research interest, its potential for technology transfer to ESnet and other providers remains limited due to the practical difficulties outlined above.

   In a recent event, a non-ESnet site user initiated a large, high-rate transfer from/to one of the ESnet sites, which was routed into ESnet via its commercial provider rather than through its REN peers. This caused the Service Level Agreement (SLA) between ESnet and its commercial provider to be violated, resulting in additional billing charges. To handle such unexpected events, the usefulness of online HNTES was considered; however, HNTES deployed within ESnet would be beneficial only for flows in one direction (leaving ESnet on to the commercial link) but not in the reverse direction. Therefore, alternate solutions such as HNTES 4.0 (end host assisted HNTES) are being explored.

7. **HNTES 3.0 GUI:** In order for network administrators to visualize information about α flows, a graphical user interface (HNTES GUI) has been designed and implemented. It displays information about α prefix flows based on parameters chosen by the user, e.g., period, router, etc. Information

about α prefix flows, such as α bytes, α time, etc. extracted from NetFlow reports will be precomputed and stored in a MySQL database. For popular user requests, e.g., the top 50 α prefix flows within one month across all routers, information will be pre-computed and stored in the database for immediate display to users. For requests involving less popular choices of parameters, e.g., all the α prefix flows observed between two specific routers in a 10-day period, GUI reports will be created in real time by issuing SQL queries, which may incur additional delays. A "drill-down" capability, as offered by other network management software such as Packet Design products, will be supported in the HNTES GUI to allow users to explore further beyond the information displayed by default. Besides these basic functions of presenting information about α prefix flows, the GUI will also offer users the ability to click on a specific α prefix flow and ask the system to initiate VC setup and router configuration for flow redirection to the newly established VC. This will require the back-end software of the HNTES GUI to have the HNTES IDC client module communicate with the OSCARS IDC for VC setup and router configuration of Layer-3 circuits.

Significant results, major findings, key outcomes and achievements:

1. **DOE ANI testbed experimental findings:**

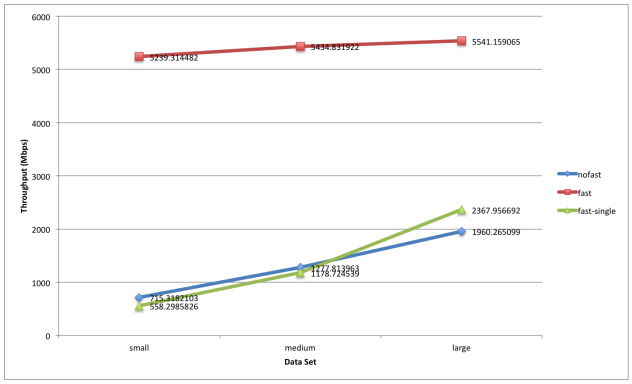

Figure 2: A comparison of GridFTP transfer throughput under different run-time settings

- The first set of experiments showed that with some arguments GridFTP transfers can achieve high rates when `-fast` is used as the data-plane TCP connections are reused for multiple file transfers. The high-rate finding implies that there is a need for isolating these flows from delay-sensitive general-purpose flows, which is the objective of HNTES. The transfer throughput was higher with `-fast` especially on high Bandwidth-Delay Product (BDP) paths because of the delay incurred in the setup/release of TCP connections for each file without `-fast`. Additionally, for each file transfer, delay was incurred due to TCP Slow Start if the `-fast` option was not used. These experiments were conducted on the 100G wide-area testbed across which the round-trip time is 48.8ms. Fig. 2 shows our results. The `small`, `medium`, and `large` indications on the x-axis correspond to three experiments in each of which an 8 GB dataset was transferred using the following settings: 128 files of size 64 MB, 32 files of size 256 MB, and 8 files of size 1 GB, respectively. The brown line (denoted `fast`) with throughput above 5000 Mbps (5 Gbps) corresponds to experiments in which the -r and -fast arguments were included. The green line (denoted `fast-single`) corresponds to experiments in which the GridFTP client, `globus-url-copy`, was initiated for each file transfer; this approach is used by some scientists who use scripts that invoke the GridFTP client separately for each file. The blue line (denoted `no-fast`) corresponds to experiments with the -r option but without the `-fast` option. This means there is a single client invocation, but the data-plane TCP connection is released after each transferred file and re-established for the next one. Finally, we learned that the Globus GridFTP implementation includes RFC 2228, FTP Security Extensions, which requires control-plane packets to be encrypted. Therefore payload based online α flow identification schemes are infeasible for any FTP application that implements this RFC.

- In the second set of experiments described above, before initiating the applications, the routers are configured to execute different QoS mechanisms. In the 1-queue case, all three flows (UDP, TCP and ping) are directed into the same output queue, emulating best-effort service. In the 2-queue case, the TCP flow packets are filtered out at the input interface, classified as α-flow packets, and directed to a separate virtual queue (called α-queue) on the output interface, while the UDP and ping flow packets are buffered in the same best-effort virtual queue. In the third configuration, 3 queues are configured, and policing is enabled on the ingress interface. Assuming that OSCARS IDC has setup a circuit of rate 1 Gbps for the TCP flow, out-of-profile packets (i.e., packets identified by the policer as exceeding this rate) are directed to a third queue called a scavenger queue. The results of the test runs are shown in Figure 3. The top graph shows the UDP flow maintaining a constant rate of 3 Gbps. The `nuttcp` TCP flow is initiated at t=53 sec. At this point, the ping delay builds up in the 1-queue configuration because the TCP flow packets arrive at high rates quickly filling up the 125 MB buffer (in an earlier experiment we quantified the size of the buffer) causing ping-generated ICMP packets to be delayed. As seen in the third plot, the average ping delay increases from 2.1 ms to over 60 ms. This emphasizes the need for isolating α-flows into their own queue. A *second* significant observation from this experiment is that in the 3-queue case, the α-flow throughput dropped significantly from 6.45 Gbps in the 2-queue case down to 0.969 Gbps in the 3-queue case (please see figure in electronic format or in a color printout to differentiate between the lines). The explanation for this finding, which was verified with additional experiments, is that TCP packets arrive out-of-sequence at the receiver when out-of-profile packets are queued in the scavenger queue, causing TCP throughput to drop. This shows the adverse effects of policing. We are now experimenting with alternative policing schemes in which the out-of-profile packets are maintained in the same α-queue, but are marked for probabilistic dropping based on buffer occupancy. These experimental findings are important for the design and prototyping of HNTES 3.0 since the QoS mechanisms need to be properly configured to meet the dual challenge of reducing the impact of α-flows on delay-sensitive flows while simultaneously not reducing their throughput.

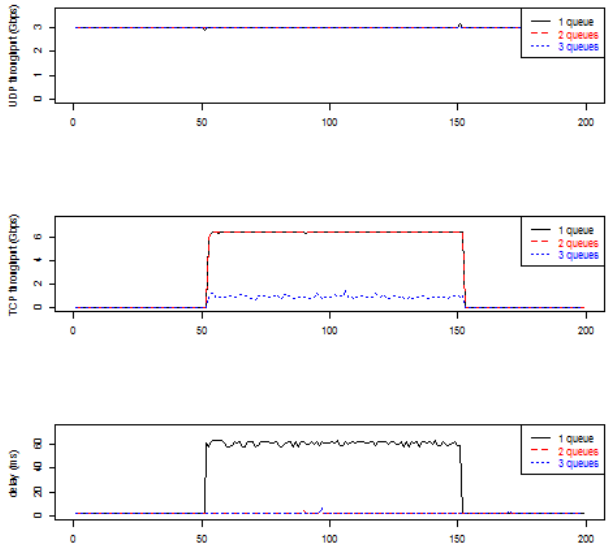

Figure 3: The impact of a high-rate TCP flow on delays illustrating the need for α-flow isolation

2. **GridFTP log analysis: (i)** Our GridFTP analysis showed that scientists do indeed move lots of files in sessions rather than single large files. Of the two datasets analyzed, NCAR-NICS data set consisted of 52,454 transfers, while SLAC-BNL dataset consisted of 1,021,999 transfers. If we set the threshold of maximum allowed gap g between transfers for inclusion within the same session to

	NCAR-NICS		SLAC-BNL	
setup delay	1 min	50 ms	1 min	50 ms
$g = 0$	0.12% (2.14%)	87.09% (89.33%)	1.95% (39.41%)	52.58% (89.68%)
$g = 1$ min	56.87% (90.54%)	92.89% (98.04%)	12.54% (78.38%)	93.56% (99.73%)
$g = 2$ min	62.16% (90.71%)	94.59% (98.17%)	15.93% (85.49%)	94.47% (99.85%)

Table 2: Percentage of sessions suitable for using VCs (percentage of transfers)

be 1 min, then there are only 211 sessions and 10,199 sessions for the two datasets, respectively. A session is considered "suitable" for dynamic VCs if the duration is at least 10 times the setup delay even under high throughput (the factor 10 is set somewhat arbitrarily). Assuming the current VC setup delay of 1 min, 56.87% of NCAR-NICS sessions, and 12.54% of SLAC-BNL sessions, would be suitable for dynamic VCs. It is interesting to point out, that these two numbers correspond to over 90% and 78% of all transfers in their respective datasets. Table 2 shows the percentage of sessions suitable for VCs for the NCAR-NICS and SLAC-BNL transfers. Furthermore, if VC setup delay can be decreased to 50 ms (e.g., with hardware implementations of signaling protocols), over 92.9% of NCAR-NICS sessions and over 93.5% of SLAC-BNL sessions would be suitable for dynamic VCs. **(ii)** The highest observed GridFTP transfer throughput is ∼4.3 Gbps. Furthermore, on all four paths for which data was analyzed, transfers occurred at rates as high as 2.5 Gbps. Thus, these transfers do consume a significant portion of link capacity, which is typically 10 Gbps on these paths, and hence these high-rate, large-sized flows have the potential for adding delays to packets of other flows. While it was previously known that scientific transfers can reach high rates, the value of this analysis lies in providing the actual observed rates. A paper on this work was accepted for publication in the Super Computing (SC2012) conference.

We analyzed the SLAC-BNL dataset, which has 1021999 transfers, to determine the percentage of transfers that use single- vs. multiple TCP flows. Of the 1021999 transfers, 864762 (84.615%) transfers consisted of multiple (more than one) parallel TCP streams (a vast majority used 8 streams).

3. **NetFlow analysis to determine the need for an online HNTES:**

We define a term *effectiveness* as the percentage of α-bytes that would have been redirected to traffic-engineered paths and isolated from the general traffic had an offline-mechanism based HNTES been deployed. If effectiveness is low, it would point to the need for an online HNTES. The per-month percentages of α-bytes that would have been redirected had the offline HNTES been deployed corresponding to different values of the aging parameter, A, are computed for the 4 routers for 7 months. The aging parameter is used to remove rules from the firewall filter configured in ingress routers. If there are no α flows corresponding to an α prefix ID in the last A aggregation intervals, the corresponding firewall filter rule is deleted. Tables 3 through 6 show the effectiveness numbers. In a separate analysis, we found that using an aging parameter of 30 is a good compromise between keeping the firewall filter tables from growing too large as new α prefix ID based rules are added, while simultaneously achieving a high effectiveness in redirecting packets from α flows. Consider the Nov. row and the column corresponding to the 30-day aging parameter in Tables 3 through 6 for all four routers. For the /24 case, the percentages are all above 85%, reaching a high value of 97.7% for the Sunnyvale core router. In all cases, the usage of /24 prefix ID is more effective than the /32 prefix ID (the negative aspect of this finding was quantified and shown to have a minimal effect presumably because HPC sites assign data transfer nodes to their own /24 subnets). With new installations of high-speed data transfer nodes, previously unseen /32 prefix identifiers would have been covered by /24 identifiers.

Next, we analyzed the NetFlow data to determine the number of **new** α prefix IDs that appeared each day. The hypothesis that α flows are created repeatedly between the same source-destination pairs is validated for the BNL analyzed data set. Overall, the trend in the new α prefix IDs graph is downward as seen with the smoothing spline function (degrees of freedom set to 4) in Fig. 4. On day 1, there were nine /24 new α prefix IDs but after day 45, in 94.7% of the days, there were only 0 or

	never	30	14	7
May	(73.1%,53.5%)	(73.1%,53.5%)	(70.6%,51.6%)	(65.6%,44.3%)
June	(95.3%,89.7%)	(94.6%,87.2%)	(93.2%,82.3%)	(88.2%,78.6%)
July	(98.9%,91.6%)	(98.2%,90.0%)	(98.1%,89.0%)	(95.6%,86.5%)
Aug	(99.3%,93.7%)	(97.7%,91.0%)	(96.0%,80.0%)	(91.3%,75.2%)
Sep	(97.3%,91.4%)	(95.0%,74.6%)	(88.9%,52.6%)	(84.6%,48.8%)
Oct	(79.6%,73.1%)	(78.4%,69.2%)	(66.2%,53.8%)	(53.8%,38.4%)
Nov	(92.0%,90.1%)	(91.5%,86.4%)	(84.4%,77.6%)	(81.2%,73.5%)

Table 3: Percentage of α bytes that would have been redirected by the end of each month for BNL (/24, /32)

	never	30	14	7
May	(48.5%,45.4%)	(48.5%,45.5%)	(48.5%,44.3%)	(48.0%,41.5%)
June	(53.2%,44.6%)	(53.2%,44.6%)	(50.1%,27.1%)	(50.1%,11.1%)
July	(58.0%,38.0%)	(56.3%,37.6%)	(52.0%,35.2%)	(46.4%,30.0%)
Aug	(85.6%,56.6%)	(81.7%,53.1%)	(76.0%,52.4%)	(63.3%,40.9%)
Sep	(87.6%,57.8%)	(70.4%,46.1%)	(63.2%,39.2%)	(51.7%,33.2%)
Oct	(88.9%,55.5%)	(79.6%,54.4%)	(71.0%,48.6%)	(66.0%,26.0%)
Nov	(90.4%,80.2%)	(85.2%,73.2%)	(81.0%,70.8%)	(75.1%,67.3%)

Table 4: Percentage of α bytes that would have been redirected by the end of each month for EQX(/24, /32)

	never	30	14	7
May	(74.5%,65.6%)	(74.5%,65.6%)	(74.4%,61.0%)	(71.9%,55.9%)
June	(94.8%,86.2%)	(93.0%,85.2%)	(91.7%,82.1%)	(89.6%,78.6%)
July	(95.0%,92.0%)	(94.5%,91.4%)	(90.1%,94.1%)	(90.0%,85.3%)
Aug	(97.3%,85.4%)	(95.0%,82.4%)	(89.7%,72.5%)	(88.1%,72.4%)
Sep	(99.0%,96.9%)	(97.7%,95.3%)	(96.0%,62.3%)	(95.6%,93.1%)
Oct	(93.4%,82.0%)	(84.6%,73.1%)	(80.1%,55.9%)	(77.8%,70.3%)
Nov	(96.8%,75.2%)	(90.2%,68.0%)	(81.5%,78.6%)	(78.0%,57.9%)

Table 5: Percentage of α bytes that would have been redirected by the end of each month for NERSC(/24, /32)

	never	30	14	7
May	(72.9%,71.5%)	(72.9%,71.5%)	(72.9%,71.5%)	(72.9%,71.5%)
June	(87.0%,82.7%)	(82.2%,77.7%)	(74.0%,72.1%)	(58.3%,56.9%)
July	(91.7%,80.6%)	(91.4%,80.5%)	(87.3%,77.3%)	(87.0%,77.3%)
Aug	(96.1%,91.7%)	(89.7%,87.9%)	(84.2%,82.3%)	(84.0%,79.7%)
Sep	(88.3%,86.2%)	(78.6%,75.9%)	(75.1%,70.1%)	(71.3%,69.6%)
Oct	(79.2%,59.9%)	(72.6%,46.6%)	(71.6%,45.5%)	(68.9%,43.3%)
Nov	(99.2%,95.4%)	(97.7%,94.6%)	(79.4%,72.0%)	(72.1%,64.8%)

Table 6: Percentage of α bytes that would have been redirected by the end of each month for SUNN(/24, /32)

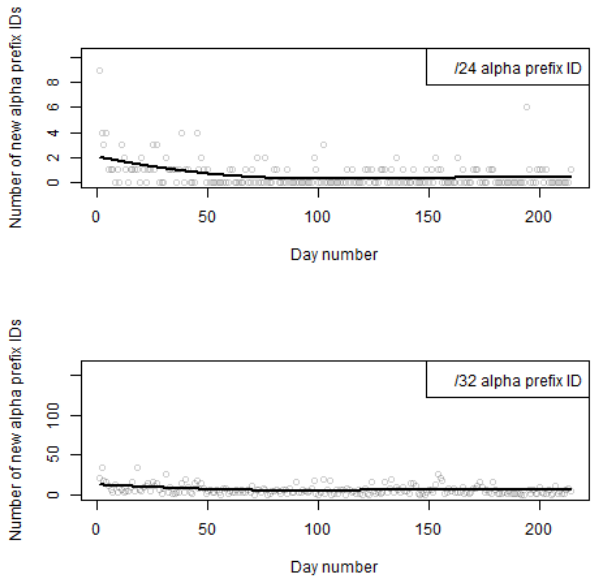

Figure 4: Number of new α prefix IDs per day with smoothing spline function (df=4) for BNL router

ESnet router	/24	/32
BNL	(125, 35)	(1548, 902)
EQX	(117, 68)	(239, 203)
NERSC	(281, 115)	(1639, 1149)
SUNN	(104, 54)	(228, 168)

Table 7: Total number of α prefix IDs observed in the 214-day period for each router, Number of α prefix IDs corresponding to which α flows appeared on only one day

1 new α prefix IDs. However, there can be days, such as Nov 10th, 2011, when α flows were observed corresponding to 6 new /24 prefix IDs, and 141 new /32 prefix IDs. This can happen when there are new installations of high-speed data transfer nodes or new scientists accessing existing data transfer nodes. These results demonstrate that while an online HNTES would help to avoid sudden unexpected traffic surges, an offline HNTES can be effective in handling α flows as the same source-destination pairs are often involved repeatedly in creating α flows.

An alternative way of visualizing the data on the number of new α flows observed each day is shown in Fig. 5 for the four routers. For each value n of the number of new α flows, N_α, the number of days in which $N_\alpha = n$ is plotted in the histograms. The total number of days is 214; therefore the y-axis can be interpreted as a probability by dividing the number of days shown by 214. All the histograms in Fig. 5 are skewed right (long tail to the right), which implies that during most days, few new α prefix flows (either /24 or /32) occurred if any. This finding, along with effectiveness values shown in Tables 3 through 6, supports a conclusion that an offline HNTES may be be sufficient for identifying, redirecting and isolating α flow packets.

Finally, Table 7 shows two numbers corresponding to each router and the two prefix ID sizes, /24 and /32. The first number in each entry within parenthesis is the total number of α prefix IDs observed in the 7-month period. The second number shows the number of α prefix IDs corresponding to which α flows appeared on just one day within the 7-month period. For the /24 subnet based data, the ratios are higher for EQX and SUNN-CR routers at 58% (i.e., 68/117), and 52%, respectively, while for the BNL and NERSC PE routers, these ratios are smaller at 28% and 40%, respectively. These numbers indicate that a significant ratio of the α prefix IDs configured into the firewall filter rules at the ingress routers will not be matched with subsequent α flows under our assumption that the HNTES offline algorithm is executed once daily. Since MPLS virtual circuits, not TDM/WDM circuits, are used to

handle α flows, and all α flows are sent to a single α queue for flow isolation from general-purpose flows, there are no wasted resources associated with these single-day α prefix IDs, except for firewall filters. But as the size of firewall filters is fairly large, it poses no direct constraint, especially under the aging parameter assumption of 30 days. A point to observe is that these ratios are for α prefix IDs, not α prefix flows. For example, there may be only one α flow in the single days in which these single-day α prefix IDs occurred in contrast to α flows that occurred corresponding to multi-day α prefix IDs. Also the bytes corresponding to these flows matters. When the results of Table 7 are interpreted in conjunction with the effectiveness ratios shown in Tables 3 through 6, it appears that the high ratios of single-day α prefix IDs may not be an issue that necessitates an online HNTES, since a large fraction of α bytes would have been redirected with the offline HNTES.

In summary, while our conclusion is not final, it appears that the data shows that a case for designing an online HNTES may be weaker than originally anticipated.

4. HNTES GUI: A preliminary GUI screenshot is shown in Fig. 6. The user can select a report period (daily, weekly, monthly), specific dates for the report period, and whether information from all routers or a specific ingress router should be included. The main window displays the router IDs of the ingress and egress routers, the source and destination addresses (anonymized; these will be replaced with organization names or subnet IP addresses by ESnet), and α bytes. The displayed screenshot is a starting point for the HNTES GUI; many more features will be added in the coming weeks.

Unexpended funds: Cost Status Show approved budget by budget period and actual costs incurred. If cost sharing is required break out by DOE share, recipient share, and total costs.

- Please see attached file.

Plans for next funding period:

- Implement the parallel TCP flow based α flow identification scheme in a new version of OFAT and transfer to Chris Tracy for execution on ESnet router NetFlow data (Dec. 14, 2012). Analyze results and compare with findings from current OFAT (completion date depends on when we obtain the anonymized results from ESnet).
- Complete HNTES 3.0 by integrating OFAT new version and IDC client (Jan. 14, 2013)
- Complete version 1 of HNTES GUI, and transfer to ESnet for testing/feedback (Mar. 1, 2013)
- Design end-user application assisted HNTES (HNTES 4.0) by running GridFTP in a shell script that issues a control-plane message to HNTES prior to starting a large session (July 15, 2013)
- Complete HNTES 4.0 prototyping for delivery to ESnet for testing (Jan. 14, 2014)

Any changes in approach or aims and reasons for change. Remember significant changes to the objectives and scope require prior approval by the contracting officer.

- None.

Actual or anticipated problems or delays and actions taken or planned to resolve them. Any absence or changes of key personnel or changes in consortium/teaming arrangement.

- None.

A description of any product produced or technology transfer activities accomplished during this reporting period, such as:

A. Publications (list journal name, volume, issue); conference papers; or other public releases of results.

- Z. Yan, C. Tracy, and M. Veeraraghavan, "A hybrid network traffic engineering system," in Proc. of IEEE 13th High Performance Switching and Routing (HPSR) 2012, June 24-27 2012.

- Z. Yan, Z. Liu, C. Tracy, and M. Veeraraghavan, "Hybrid network traffic engineering system (HNTES)," ESCC Meeting at Joint Techs, Baton Rouge, LA, Jan. 25-26, 2012.
- Z. Yan, Z. Liu, C. Tracy, and M. Veeraraghavan, "Hybrid Network Traffic Engineering System," Talk presented at Annual DOE PI meeting for the ASCR Network & Middleware, Mar. 1-2, 2012.
- Z. Yan, C. Tracy, and M. Veeraraghavan, "Hybrid network traffic engineering system (HNTES)," Internet2 Spring Member Meeting, April 2012.
- Z. Liu, M. Veeraraghavan, Z. Yan, C. Tracy, J. Tie, I. Foster, J. Dennis, J. Hick, Y. Li and W. Yang, "On using virtual circuits for GridFTP transfers," SuperComputing (SC2012), Salt Lake City, Nov. 10-16, 2012.

B. Web site or other Internet sites that reflect the results of this project.

- UVA Hybrid Networking Project web site: http://www.ece.virginia.edu/mv/research/DOE09/index.html
- Collaboration web site for project participants to post documents, discuss issues in an online forum, archive emails, etc. The password-protected site is: https://collab.itc.virginia.edu/portal/site/e121f110-7b37-4021-8ac1-4d61197c067a/page/d4fece61-b037-411e-9866-c8a890ee22c2

C. Networks or collaborations fostered.

- We are working closely with the Chris Tracy, ESnet, and the ANI Tabletop Testbed design team, including Brian Tierney, Inder Monga, Chin Guok and Eric Pouyoul, ESNet. For the GridFTP work, we established collaborations with Ian Foster, Raj Kettimuthu, and Linda Winkler, ANL, Brent Draney, Jason Hick, Jason Lee, NERSC, and Yee-Ting Li and Wei Yang, SLAC.

D. Technologies/Techniques.

- Our algorithms are part of documents posted on the project Web site under Documents.

E. Inventions/Patent Applications

- None.

F. Other products, such as data or databases, physical collections, audio or video, software or netware, models, educational aid or curricula, instruments or equipment.

- The software developed by our team is available on our project Web site..

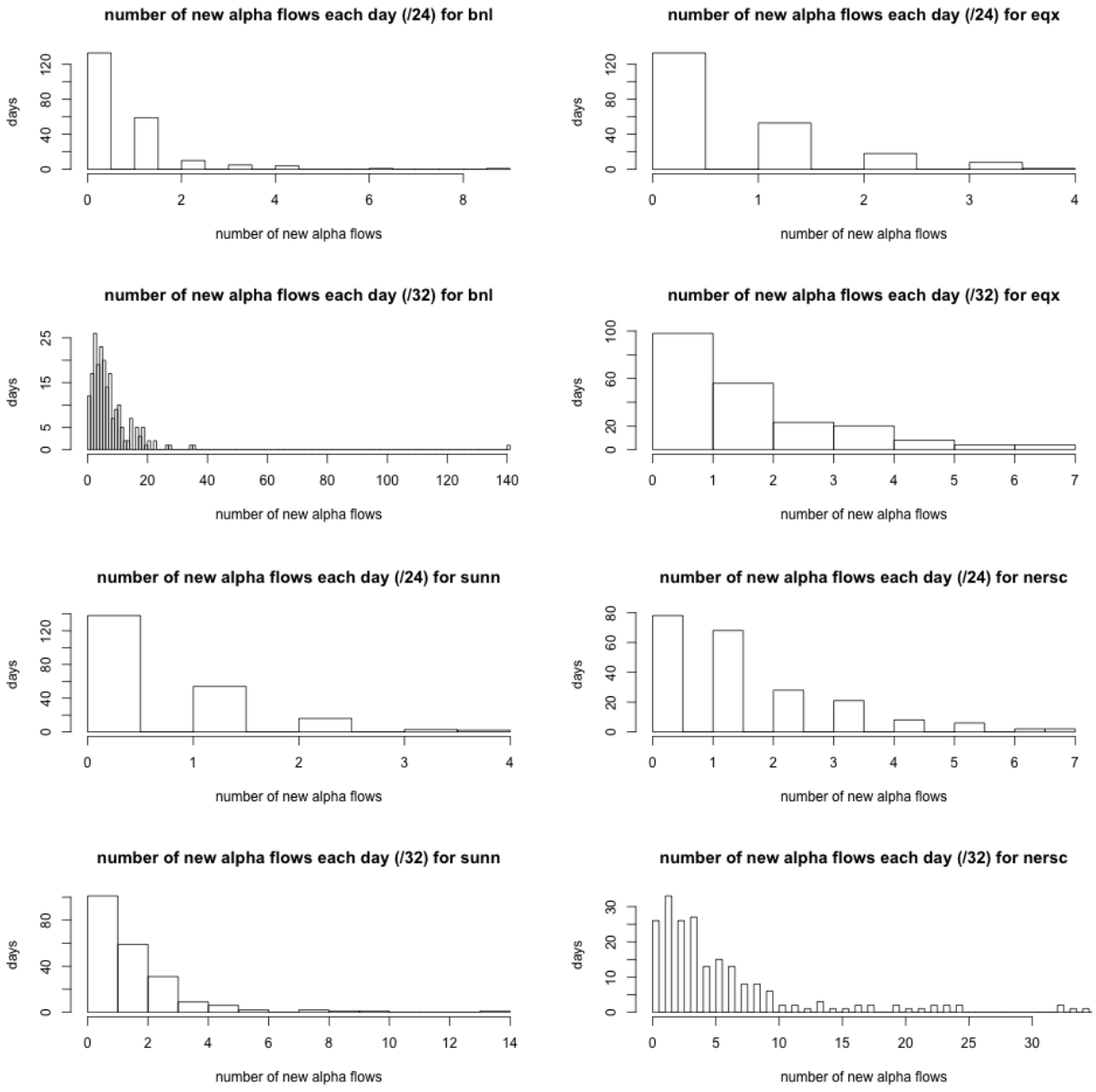

Figure 5: Histograms for the number of new α flows identified in each day for 4 routers

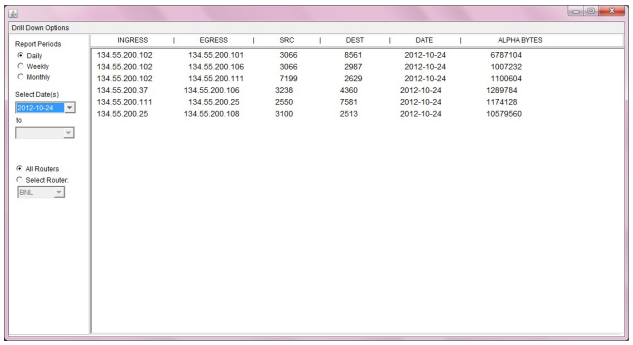

Figure 6: An example display of the HNTES GUI